

International Conference
APPLIED STATISTICS 2022
Abstracts and Program



September 19–21, 2022
Ljubljana, Slovenia

International Conference
APPLIED STATISTICS 2022
Abstracts and Program

September 19–21, 2022
Ljubljana, Slovenia

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 120992003
ISBN 978-961-94283-2-0 (PDF)

International Conference
APPLIED STATISTICS 2022
Abstracts and Program

Scientific Program Committee

Lara Lusa (Chair), Slovenia
Mihael Perman (Scientific advisor), Slovenia
Janez Stare (Scientific advisor), Slovenia
Vladimir Batagelj, Slovenia
Andrej Blejec, Slovenia
Matevž Bren, Slovenia
Maurizio Brizzi, Italy
Anuška Ferligoj, Slovenia
Herwig Friedl, Austria
Dario Gregori, Italy
Katarina Košmelj, Slovenia
Irena Križman, Slovenia
Stanislaw Mejza, Poland
Jože Rován, Slovenia
Tamas Rudas, Hungary
Vasja Vehovar, Slovenia

Organizing Committee

Irena Vipavc Brvar (Chair)
Jerneja Čuk
Bogdan Grmek
Andrej Kastrin
Lara Lusa

<i>Edited by</i>	Andrej Kastrin and Lara Lusa
<i>Published by</i>	Statistical Society of Slovenia Litostrojska c. 54 1000 Ljubljana, Slovenia
<i>Publication year</i>	2022
<i>Publication format</i>	PDF
<i>Electronic version:</i>	https://akastrin.si/as/as-book-2022.pdf

ABSTRACTS and PROGRAM

PROGRAM

		Room 1	Room 2
Monday	9.00 – 9.10	Opening	
	9.10 – 10.00	Invited lecture	
	10.00 – 10.30	Break	
	10.30 – 12.00	Biostatistics	
	12.00 – 14.00	Break	
	14.00 – 15.15	Biostatistical applications	
	15.15 – 15.30	Break	
	15.30 – 17.00	Invited session (Blockmodeling)	
Tuesday	9.10 – 10.00	Invited lecture	
	10.00 – 10.30	Break	
	10.30 – 12.00	Modeling & simulation	
	12.00 – 14.00	Break	
	14.00 – 15.15	Network analysis	
	15.15 – 15.30	Break	
	15.30 – 17.00	Education and social sciences	
	17.00 – 17.15	Break	
	17.15 – 18.05	Invited lecture	
	18.05 – 18.15	Closing	
Wednesday	9.00 – 12.00	Workshop	

9.00–9.10 **Opening**
Room 1

9.10–10.00 **Invited lecture**
Room 1

Chair: Maja Pohar Perme

1. **Competing risks methods for discrete time-to-event data**
Matthias Schmid

10.00–10.30 **Break**

10.30–12.00 **Biostatistics**
Room 1

Chair: Matthias Schmid

1. **Use of data analytics to enhance clinical research and improve patient engagement and experience**
Jay Mandrekar
2. **Analysing transition probability in non-Markov multi-state models**
Maja Pohar Perme, Per Kragh Andersen and Eva Nina Sparre Wandall
3. **Anamnestic descriptors of chronic pain development: A multivariate analysis**
Gaj Vidmar and Helena Jamnik
4. **How is predictive modeling in biomedicine changing and is there a need for new guidance principles?**
Lara Lusa and Joerg Rahnenfuehrer
5. **Estimating the life years lost of cancer patients**
Damjan Manevski, Maja Pohar Perme and Tina Košuta
6. **Evaluating cancer screening programmes using survival analysis**
Bor Vratinar and Maja Pohar Perme

12.00–14.00 **Break**

14.00–15.15 **Biostatistical applications**
Room 1

Chair: Damjan Manevski

1. **Statistical analysis of EMF triggered symptoms within electromagnetic hypersensitive population**
Maurizio Brizzi and Fiorenzo Marinelli

2. Modelling vaccination and virus variant effects in a cohort of patients hospitalized for acute COVID-19

Nataša Kejžar and Daša Stupica

3. SARS-CoV-2 vaccine effectiveness against different variants of concern: A meta-analysis of test-negative design case-control studies

Sara Tagliaferri, Matilde Passerini, Paola Rabatelli, Francesco Capriotti, Giuseppe Pedrazzi and Carlo Ferrari

4. Mortality, seasonal variation, and susceptibility to acute exacerbation of COPD in the pandemic year: A nationwide population study

Irena Šarc, Aleša Lotrič Dolinar, Tina Morgan, Jože Sambt, Kristina Zihel, Dalibor Gavrić, Julij Šelb, Aleš Rozman and Petra Došenović Bonča

5. The types of online social support sought by women, victims of violence

Vanja Erčulj

15.15–15.30 **Break**

15.30–17.00 **Invited session (Blockmodeling)**

Room 1

Chair: Aleš Žiberna

1. Model-based clustering in hypergraphs through a stochastic blockmodel

Catherine Matias and Luca Brusa

2. Learning common structures in a collection of networks

Saint-Clair Chabert-Liddell

3. Approaches for blockmodeling temporal networks

Marjan Cugmas and Aleš Žiberna

9.10–10.00 **Invited lecture**
Room 1

Chair: Mihael Perman

1. **Robust linear and logistic regression for high-dimensional compositional data**
Peter Filzmoser

10.00–10.30 **Break**

10.30–12.00 **Modeling & simulation**
Room 1

Chair: Peter Filzmoser

1. **Modeling COVID-19 epidemic in Slovenia**
Janez Žibert, Miha Fošnarič and Tina Kamenšek
2. **GOFLMM.jl: Goodness-of-fit for linear mixed models**
Jakob Peterlin
3. **A simulation study of the limitations of Bland Altman analysis**
Maša Kušar
4. **Measure of expected goals in football and its application**
Urh Peček
5. **Bayesian estimation for linear regression model with multicollinearity problem using Gibbs sampling: A simulation study**
Monthira Duangsaphon, Teerawat Simmachan, Kamon Budsaba and Rattana Lerdsuwansri
6. **Determining the changepoints in segmented regression**
Ana Radulović and Rok Blagus

12.00–14.00 **Break**

14.00–15.15 **Network analysis**
Room 1

Chair: Marjan Cugmas

1. **Projections of weighted two-mode networks**
Vladimir Batagelj
2. **Scaling limits for parking on Frozen Erdős-Rényi Cayley trees with heavy tails**
Andrej Srakar
3. **Stochastic generalized blockmodeling**
Aleš Žiberna

4. **Making blockmodeling easier with BlockmodelingGUI: R Shiny as a GUI for generalised block-modelling**

Fabio Ashtar Telarico

5. **Authors and research topics presented at the Applied Statistics conference (2010–2022)**

Polona-Maja Repar, Bor Vratinar and Andrej Kastrin

15.15–15.30 **Break**

15.30–17.00 **Education and social sciences**

Room 1

Chair: Nataša Kejžar

1. **Brownian excursions and the Kolmogorov statistic**

Mihael Perman and Jon A. Wellner

2. **Do graduate students of business and management view statistics and research skills as relevant for their careers?**

Irena Ograjenšek and Iddo Gal

3. **Science literacy and enjoyment in learning science: A statistical perspective**

Melita Hajdinjak

4. **Online survey panels in modern social science research: Can non-probability sampling replace probability sampling?**

Gregor Čehovin and Vasja Vehovar

5. **Why logistic regression is not always suitable for risk ratio estimation**

Jon Pustavrh-Mičović and Nikolaj Candellari

6. **Wage gap between men and women in the Israeli economy**

Tal Shahor and Javier Simonovich

17.00–17.15 **Break**

17.15–18.05 **Invited lecture**

Room 1

Chair: Vasja Vehovar

1. **Applications of multiple imputation, from calibrating changing occupation/industry coding systems in the US Census, to evaluating the impact of nonresponse in the Slovenian Plebiscite, to understanding placebo effects in pharmaceutical experiments**

Donald B. Rubin

18.05–18.15 **Closing**

Room 1

9.00–12.00 **Workshop**
Room 2

1. **Generalized blockmodeling in R using blockmodeling package**
Aleš Žiberna and Marjan Cugmas

ABSTRACTS

Invited lecture

Competing risks methods for discrete time-to-event data

Matthias Schmid

University of Bonn, Bonn, Germany
matthias.c.schmid@uni-bonn.de

This talk presents an overview of statistical methods for the analysis of discrete failure times with competing events. We describe the most commonly used modeling approaches for this type of data, including discrete versions of the cause-specific hazards model and the subdistribution hazard model. In addition to discussing the characteristics of these methods, we present approaches to nonparametric estimation and model validation. Our literature review suggests that discrete competing-risks analysis has gained substantial interest in the research community and is used regularly in econometrics, biostatistics, and educational research.

Biostatistics

Use of data analytics to enhance clinical research and improve patient engagement and experience

Jay Mandrekar

Mayo Clinic, Rochester, United States
mandrekar.jay@mayo.edu

Research in academic medical centers offer opportunities to collaborate on clinical projects that require novel application of both common and uncommon statistical methods. This talk will focus on two aspects. First part will discuss some of the challenges encountered while conducting clinical research. These will include setting up databases, recruitment of participants, missing data issues on key variables of interest. Some examples, on overcoming some of these challenges using novel data analytic techniques will be discussed from a Data Scientist's perspective. Second part will illustrate how one can use novel data analytic techniques such as exploratory factor analysis and logistic regression for improving patient participation, engagement and experience in clinical research studies.

Analysing transition probability in non-Markov multi-state models

Maja Pohar Perme¹, Per Kragh Andersen² and Eva Nina Sparre
Wandall²

¹University of Ljubljana, Ljubljana, Slovenia

²University of Copenhagen, Copenhagen, Denmark

maja.pohar@mf.uni-lj.si, pka@biostat.ku.dk, eva.sparre.wandall@sund.ku.dk

Multi-state models are frequently used when data come from subjects observed over time and where focus is on the occurrence of events that the subjects may experience. A convenient modeling assumption is that the multi-state stochastic process is Markovian, in which case a number of methods are available when doing inference for both transition intensities and transition probabilities. The Markov assumption, however, is quite strict and may not fit actual data in a satisfactory way. Therefore, inference methods for non-Markov models are needed. In this talk, we address the problem of estimating transition probabilities in such models and suggest ways of doing regression analysis based on pseudo observations. In particular, we will compare methods using landmarking with methods using plug-in. The methods are illustrated using simulations.

Anamnestic descriptors of chronic pain development: A multivariate analysis

Gaj Vidmar and Helena Jamnik

University Rehabilitation Institute, Ljubljana, Slovenia
gaj.vidmar@ir-rs.si, helena.jamnik@ir-rs.si

There is little knowledge about possible trajectories of chronic pain development. By analysing patients' medical records and other available clinical data, we attempted to identify specific development patterns of pain and other symptoms in patients who attended a specialist outpatient rehabilitation clinic for chronic non-malignant pain. We included 487 consecutive adult patients and used 16 binary descriptors to characterise symptom development. We first estimated the most common descriptors (sleep disorders and physical or mental fatigue during chronic pain development, gradual development from local, usually low-back, to widespread pain) and examined the correlations between the descriptors (using phi coefficient). Hierarchical clustering based on the descriptors (using Euclidean distance and Ward's method) identified three distinct groups of patients; the fourth group were those with an unclear medical history. In addition, we explored the relations between the descriptors among the 430 patients with a clear medical history using multidimensional scaling (in two dimensions, based on squared Euclidean distance and the PROXSCAL algorithm) and multiple correspondence analysis (in two dimensions).

How is predictive modeling in biomedicine changing and is there a need for new guidance principles?

Lara Lusa¹ and Joerg Rahnenfuehrer²

¹University of Primorska/University of Ljubljana, Koper (Capodistria)/Ljubljana, Slovenia

²TU Dortmund University, Dortmund, Germany

lara.lusa@mf.uni-lj.si, rahnenfuehrer@statistik.tu-dortmund.de

The number of predictive models proposed in the biomedical literature is growing every year. In the last few years there has been an increasing attention to the changes that are occurring in the predictive modeling landscape, which are mostly related to the methods and data that are being used. For example, in predictive modelling related to human diseases, machine learning techniques are becoming more and more popular, supported by some spectacular results of deep learning techniques in applications with large numbers of observations. Also, the models are often developed using complex data such as images, electronic health records, registries, genomic and biomarker data. Many influential editorials suggested that the existing best practice recommendations for design, conduct, analysis, reporting, impact assessment, and clinical implementation from the traditional biostatistics and medical statistics literature, are not sufficient to guide the use of predictive models using new (machine learning/artificial intelligence) methods and complex data. In this talk we will address two points: (i) how is predictive modelling changing in practice and (ii) what are the (specific) needed recommendations for predictive modeling. We base our findings on the evidence emerging from systematic reviews and the analysis of guideline documents published in the last few years. We address also the differences between machine learning and statistical models.

Estimating the life years lost of cancer patients

Damjan Manevski, Maja Pohar Perme and Tina Košuta

University of Ljubljana, Ljubljana, Slovenia

damjan.manevski@mf.uni-lj.si, maja.pohar@mf.uni-lj.si, tina.kosuta@mf.uni-lj.si

When evaluating the long-term survival of a given patient cohort we may estimate the life years lost due to the studied disease as a suitable summary measure. This allows for a comparison of the cohort survival to the expected survival in the general population and thus one can provide new possibly interesting insight into the survival of patients with the given disease. Different measures of life years difference may be considered in practice. In this talk, we will use the example of breast cancer data provided by the Slovene cancer registry. A suitable measure will be chosen for calculating the life years difference compared to the general population. The focus will be on non-parametric estimation and some of the estimation difficulties will be considered. For each individual, estimating the expected population curve is straightforward, whereas the estimation of his observed curve is a more complex challenge. There are two basic problems: first, the two time scales (age and time since diagnosis) may affect the time to event, thus the Markov assumption (which is assumed in the model) may be violated in which case both time scales have to be considered in the analysis. Second, patients enter the data set at different times (in our example, patients are diagnosed in a time span of 70 years). Consequently, the follow-up times differ across patients and informative censoring due to the date of entry can cause biased estimates. We will consider these challenges using the given data set and consider some of the possible improvements. The practical use of this work will be also illustrated using the R package `reلسurv`.

Evaluating cancer screening programmes using survival analysis

Bor Vratinar and Maja Pohar Perme

University of Ljubljana, Ljubljana, Slovenia

bor.vratinar@mf.uni-lj.si, maja.pohar@mf.uni-lj.si

Cancer screening is a programme for medical screening of asymptomatic people who are at risk of developing cancer. Typically, participants are regularly screened every few years using blood tests, urine tests, medical imaging, or other methods. Among cases who are screened regularly some are diagnosed with cancer based on screening tests (screen-detected cases) and some based on symptoms appearing in the interval between two consecutive screening tests (interval cases). The hypothesis is that the screening programmes improve chances of survival for screen-detected cases as these cases are diagnosed and treated at an earlier stage of the disease compared to counterfactual scenario where their cancer would have been detected based on symptoms. We would like to test this hypothesis empirically. So far, the problem has been tackled by comparing the survival functions of screen-detected cases and interval cases. Realizing that the direct comparison between these two groups would result in biased results, previous research focused on parametric solutions to remove the bias. We argue that the problem lies elsewhere—that this comparison, in fact, does not reflect the question of interest. Therefore, in this study, we precisely define the contrast corresponding to the hypothesis defined above. Since the contrast of interest refers to hypothetical quantities, we discuss which data and under what assumptions can be used for estimation. We also propose a non-parametric framework for evaluating the effectiveness of cancer screening programmes under certain assumptions. The proposed ideas are illustrated using simulated data. The problem is motivated by the need to evaluate breast cancer screening programme in Slovenia.

Biostatistical applications

Statistical analysis of EMF triggered symptoms within electromagnetic hypersensitive population

Maurizio Brizzi¹ and Fiorenzo Marinelli²

¹University of Bologna, Bologna, Italy

²Institute of Molecular Genetics, National Research Council (CNR), Bologna, Italy

maurizio.brizzi@unibo.it, fmarinel@area.bo.cnr.it

Electromagnetic hypersensitivity (EHS) is a health condition characterized by a variety of non-specific symptoms attributed by the affected individuals to the exposure to EMF. The symptoms most commonly reported include dermatological symptoms (redness, tingling, and burning sensations), neurological symptoms (fatigue, tiredness, concentration difficulties, dizziness) and vegetative alterations (nausea, heart palpitation and digestive disturbances). The symptoms often occur after an electromagnetic exposure and they usually disappear or improve with the removal of the exposure source. Due to the biological mediation of the electromagnetic exposure, an important question is to detect the causal correlation between exposure and symptoms. We used a personal dosimeter (ESM-140 MASCHECK) to assess the exposures to RF EMF of 25 patients with self-reported EHS for approximately 24 hours, considering 8 different kinds of exposure. The patients were asked to keep a diary of the symptoms they attribute to EMF exposure. We have then searched for a possible statistical correlation between the exposure peaks, detected by the dosimeter, and the onset of symptoms reported on the diary. We have divided the assessment period in time slots of 48 minutes each, thus having 30 slots in 24 hours: for each patient and slot, we have recorded the maximum exposure to EMF waves and the number of symptoms suffered by the patient. The first observed results suggest the presence of a causal relationship between exposure to EMF and symptoms, since it seems that symptoms come much more frequently when the exposure is relevant. This correspondence is particularly evident with exposure to GSM and WIFI frequencies. Grant: Association on environmental and chronic toxic injury (A.M.I.C.A.), Rome, Italy

Modelling vaccination and virus variant effects in a cohort of patients hospitalized for acute COVID-19

Nataša Kejžar¹ and Daša Stupica^{1,2}

¹University of Ljubljana, Ljubljana, Slovenia

²University Medical Centre Ljubljana, Ljubljana, Slovenia

natasa.kejzar@mf.uni-lj.si, dasa.stupica@kclj.si

We collected data about the cohort of adult patients (≥ 18 years) hospitalised in autumn 2021 and spring 2022, where the Delta and Omicron SARS-CoV-2 variants were predominant. The data were used to evaluate the association between (i) primary vaccination and (ii) Delta/Omicron virus variant and the primary outcome—progression to critically severe disease (mechanical ventilation or death). Descriptive group comparison showed that fully vaccinated patients, and patients hospitalized in the period of Omicron predominance were older, more often immunocompromised, and had higher Charlson comorbidity index scores. We analysed the primary outcome (critically severe disease) using logistic regression model (adjusted for selected covariates). Secondary outcome—the time of discharge/death—was analysed by multi-state and Fine-Gray time-to-event models. All analyses were complemented with the analyses of propensity-score matched sample. We present the data and models with their interpretation and discuss potential issues regarding model fit as well as effect interpretation.

SARS-CoV-2 vaccine effectiveness against different variants of concern: A meta-analysis of test-negative design case-control studies

Sara Tagliaferri¹, Matilde Passerini², Paola Rabatelli², Francesco Capriotti², Giuseppe Pedrazzi² and Carlo Ferrari²

¹University of Parma (Department of Medicine and Surgery / CERT, Center of Excellence for Toxicological Research), Parma, Italy

²University of Parma, Department of Medicine and Surgery, Parma, Italy

sara.tagliaferri@unipr.it, matilde.passerini@studenti.unipr.it,
paola.rabatelli@unipr.it, francesco.capriotti@unipr.it, giuseppe.pedrazzi@unipr.it,
carlo.ferrari@unipr.it

Efficacy of COVID-19 vaccine has been demonstrated by phase 3 randomized-controlled trials. However, clinical trials did not address important questions regarding vaccine performance in real-world conditions. The aim of the current study was to systematically retrieve and critically appraise available evidence on effectiveness of mRNA vaccines against recent SARS-CoV-2 variants in preventing infection, hospitalization and severe disease. We conducted a systematic search on PubMed and by checking reference lists. We screened titles and abstracts, and reviewed full texts to identify relevant articles. Independent reviewers selected observational studies with the outcome of interest, extracted data, and assessed the risk of bias. Meta-analysis was performed including studies with a test-negative case-control design, where raw data on proportion of infected, hospitalized or with severe disease fully vaccinated (2 doses and booster) and unvaccinated subjects were available to calculate the odds ratio (OR), as effect measure. The DerSimonian-Laird random- and fixed-effects models were used. 19 studies were included in the review, showing that vaccine effectiveness (VE) against infection, hospitalization and severe disease was higher for Delta when compared to Omicron. Meta-analysis was performed only for 10 studies reporting infection and hospitalization data. The subgroup analysis (Omicron and Delta) showed a significant protection of mRNA vaccines against infection, both for Omicron and Delta (OR 2.19, IC95% 1.62-2.97; OR 9.85, IC95% 2.38-40.75, respectively), with a significant difference between them ($p=0.04$). The proportion of hospitalized SARS-Cov-2 cases was higher among fully vaccinated versus unvaccinated subjects infected by both variants (OR 3.49, IC95% 1.79-6.80; OR 17.08, IC95% 6.18-47.24, respectively). VE against hospitalization was significantly higher for Delta when compared to Omicron ($p=0.01$). The duration of effectiveness of current vaccines is preserved against infection and hospitalization induced by the emerging variants. Furthermore,

the VE against infection and hospitalization tends to be reduced for Omicron compared to Delta.

Mortality, seasonal variation, and susceptibility to acute exacerbation of COPD in the pandemic year: A nationwide population study

Irena Šarc¹, Aleša Lotrič Dolinar², Tina Morgan¹, Jože Sambt²,
Kristina Zihherl¹, Dalibor Gavrić³, Julij Šelb¹, Aleš Rozman¹ and
Petra Došenović Bonča²

¹University Clinic of Respiratory and Allergic Diseases, Golnik, Slovenia

²University of Ljubljana, Ljubljana, Slovenia

³Health Insurance Institute of Slovenia, Ljubljana, Slovenia

irena.sarc@klinika-golnik.si, ales.lotric.dolinar@ef.uni-lj.si,
tina.morgan@klinika-golnik.si, joze.sambt@ef.uni-lj.si,
kristina.zihherl@klinika-golnik.si, dalibor.gavric@zzzs.si,
julij.selb@klinika-golnik.si, ales.rozman@klinika-golnik.si,
petra.d.bonca@ef.uni-lj.si

Previous studies have suggested that the coronavirus disease 2019 (COVID-19) pandemic was associated with a decreased rate of acute exacerbation of chronic obstructive pulmonary disease (AECOPD), however the data on how the COVID-19 pandemic has influenced mortality, seasonality of, and susceptibility to AECOPD in the COPD population was scarce. We conducted a national population-based retrospective study using data from the Health Insurance Institute of Slovenia from 2015 to February 2021, with 2015–2019 as the reference period. We extracted patient and hospitalisation data for AECOPD. The national COPD population was generated based on dispensed prescriptions of inhalation therapies, and moderate AECOPD events in the population were analysed based on dispensed AECOPD medications. The numbers of severe and moderate AECOPD were reduced by 48% and 34%, respectively, in the pandemic year 2020. In 2020, the seasonality of AECOPD was reversed, with a 1.5-fold higher number of severe AECOPD in summer compared to winter. The proportion of frequent exacerbators (≥ 2 AECOPD hospitalisations per year) was reduced by 9% in 2020, with a 30% reduction in repeated severe AECOPD in frequent exacerbators and a 34% reduction in persistent frequent exacerbators (≥ 2 AECOPD hospitalisations per year for 2 consecutive years) from 2019. The number of patients with two or more moderate AECOPD decreased by 40% in 2020. In 2020, non-COVID mortality decreased (-15.3%) and no excessive mortality was observed in the COPD population. In the pandemic year, we found decreased susceptibility to AECOPD across severity spectrum of COPD, reversed seasonal distribution of severe AECOPD and decreased non-COVID mortality in the COPD population, potentially reflecting the effect of reduced respiratory virus circulation due to public health measures.

The types of online social support sought by women, victims of violence

Vanja Erčulj

University of Maribor, Ljubljana, Slovenia
vanja.erculj@um.si

One in three women experience an intimate partner or other type of violence. Social support has an important role in alleviating stress of the support seeker and enhance the search for possible solutions of the problem at hand. Seeking online social support might be within easier reach for women, victims of intimate partner or other type of violence. Online social support was shown to have an important role as it encourages women to leave their abusive partners. The objective of this research was to explore what types of social support women seek online and what kind of abusive situations drive them to seek online social support. For this purpose, all the posts from 2002 to the beginning of 2020 were retrieved from the Slovenian online social support group Women in need, the first such online social support in Slovenia. The analysis included manual annotation of 600 randomly chosen beginning posts of the users. The type of social support and experienced violence as well as the actor of the violence were annotated by two independent annotators and the discrepancies were addressed by the third annotator. Logistic regression was used to explore the association between the types of social support women sought and experiencing different types of violence. Text clustering was used to identify the main needs women expressed in their beginning posts to the online support group. The results suggest that women mainly seek informational support and that certain types of experienced violence are to higher extent associated with the type of social support sought. The content of the beginning posts expressed the main difficulties and uncertainties regarding official procedures necessary for the improvement of the situation and verbal aggression and alienation encountered in the partner relationship.

Invited session (Blockmodeling)

Model-based clustering in hypergraphs through a stochastic blockmodel

Catherine Matias¹ and Luca Brusa²

¹CNRS - Sorbonne Universite, Paris, France

²University of Milano Bicocca, Milano, Italia

catherine.matias@math.cnrs.fr, luca.brusa@unimib.it

Over the past few decades, a broad variety of models has been developed for graphs. However, modern applications in various fields highlighted the need to account for higher-order interactions, to include information deriving from groups of three or more nodes. Simple examples include group interactions in social networks, scientific co-authorship, interactions between more than two species in ecological models or high-order correlations between neurons in brain networks. Hypergraphs provide the most general formalization of higher-order interactions: similarly to a graph, a hypergraph is defined as a set of nodes and a set of hyperedges, the latter specifying nodes taking part in each interaction. We propose a stochastic block model for hypergraphs to perform model-based clustering, capturing the information deriving from higher-order interactions. A discrete latent variable with Q support points is associated to each node, identifying the latent states in the population. The model parameters are the weight of each latent state, and the occurrence probability of a hyperedge given the belonging latent states of its nodes. The formulation of the model is sufficiently flexible to account for possible simplified latent structures; an example is the situation in which the conditional probability of occurrence of an hyperedge can only assume two possible values: one if all its nodes belong to the same latent state, and the other otherwise. Maximum likelihood estimation of model parameters is performed through a variational expectation-maximization algorithm, by maximizing a lower bound of the log-likelihood function. Spectral clustering techniques are employed to provide an optimal initialization to the algorithm, and model selection is explored using the ICL criterion. The model is applied to both simulated and real data, and the performance of the proposal is assessed in terms of parameter estimation and ability to recover the clusters (through the Adjusted Rand Index). The estimation algorithm is implemented in C++ language (both in serial and in parallel version) and it is made available for the R software.

Learning common structures in a collection of networks

Saint-Clair Chabert-Liddell

Agroparistech Innovation, Paris, France
academic@chabert-liddell.com

Let a collection of networks represent interactions within several (social or ecological) systems. Two main issues arise: identifying similarities between the topological structures of the networks or clustering the networks according to the similarities in their structures. We tackle these two questions with a probabilistic model based approach. We propose an extension of the Stochastic Block Model (SBM) adapted to the joint modeling of a collection of networks. The networks in the collection are assumed to be independent realizations of SBMs. The common connectivity structure is imposed through the equality of some parameters. The model parameters are estimated with a variational Expectation-Maximization (EM) algorithm. We derive an ad-hoc penalized likelihood criterion to select the number of blocks and to assess the adequacy of the consensus found between the structures of the different networks. This same criterion can also be used to cluster networks on the basis of their connectivity structure. It thus provides a partition of the collection into subsets of structurally homogeneous networks. We apply our proposition on two collections of networks. First, an application to advice networks between judges, lawyers, priests and researchers reveals which networks share common structures as well as the correspondence between groups of actors in different systems playing equivalent social roles. We also show how using the information contained in networks sharing similar structures improves the prediction of missing data. Second, we cluster 67 food webs according to their connectivity structures and demonstrate that five mesoscale structures are sufficient to describe this collection.

Approaches for blockmodeling temporal networks

Marjan Cugmas and Aleš Žiberna

University of Ljubljana, Ljubljana, Slovenia

marjan.cugmas@fdv.uni-lj.si, ales.ziberna@fdv.uni-lj.si

Blockmodeling refers to a set of approaches for simplifying complex network structures. Blockmodeling approaches for networks observed at one time point are already well developed and widely used. However, this is not true for blockmodeling temporal networks, for which approaches have only recently been introduced. Several introduced approaches for temporal networks have the same goal (i.e., to find a partition of equivalent nodes considering their temporal dependency), but they differ greatly in how they achieve the goal, including the way they account for temporal dependency. Since these approaches are new, they are not yet widely used and have not yet been compared through simulations. Therefore, it is not known in which cases a practitioner should use blockmodeling for temporal networks versus “regular” blockmodeling, how sensitive the block modelling approaches for temporal networks are to different network characteristics, and which blockmodeling approach for temporal networks should be preferred. The above questions are addressed in this presentation. Different blockmodeling approaches were analysed using Monte Carlo simulations, generating networks with different characteristics. Special attention has been paid to generating networks considering local network mechanisms, making the generated networks more similar to real social networks. The other considered network characteristics are network size, block densities, blockmodel type change, and stability of clusters in time. The results suggest that separate analysis of networks at different time points is sufficient in some cases. However, the use of blockmodeling approaches for temporal networks may be beneficial when there is some dependence between partitions from successive time points. The DSBM (Matias and Miele, Royal Society Open Science, 2017, 4(6), 1–10) approach is most efficient when a blockmodel type does not change, and SBMfMLN (Bar-Hen et al., Statistical Modelling, 2020, 1–24) when it does.

Invited lecture

Robust linear and logistic regression for high-dimensional compositional data

Peter Filzmoser

Vienna University of Technology, Vienna, Austria

p.filzmoser@tuwien.ac.at

Compositional data analysis (CoDa) is based on analyzing log-ratios between the variables, and it is very useful in contexts where the measured values themselves are not meaningful, and thus the variables need to be compared relative to each other. An example are microbiome data, where the relevant information is contained in relative bacterial taxa abundances rather than in the absolute ones. Such data are typically high-dimensional, and usually only a small subset of bacteria is related to an external property. Another frequent problem with real data are outliers or observations which are inconsistent for some reason. Such observations could negatively affect the parameter estimation, with the consequence that the model is no longer appropriate, neither for regular observations nor for the outliers. We focus on both problems and introduce an elastic-net penalized estimator for linear as well as for logistic regression with compositional data. The proposed methods are based on the so-called log-contrast model, and robustness is achieved by trimming the objective function. We show the advantages of the resulting methods for simulated and for real microbiome data. R code has been made available at <https://github.com/giannamonti/RobZS>.

Modeling & simulation

Modeling COVID-19 epidemic in Slovenia

Janez Žibert, Miha Fošnarič and Tina Kamensek

University of Ljubljana, Ljubljana, Slovenia

janez.zibert@zf.uni-lj.si, miha.fosnarič@zf.uni-lj.si, tina.kamensek@zf.uni-lj.si

In the absence of a systematic approach to epidemiological modeling in Slovenia, various isolated mathematical epidemiological models emerged shortly after the outbreak of the COVID-19 epidemic. We present an epidemiological model adapted to the COVID-19 situation in Slovenia. The standard SEIR model was extended to distinguish between age groups, symptomatic or asymptomatic disease progression, and vaccinated or unvaccinated populations. Evaluation of the model forecasts for 2021 showed the expected behavior of epidemiological modeling: our model adequately predicts the situation up to 4 weeks in advance; the changes in epidemiologic dynamics due to the emergence of a new viral variant in the population or the introduction of new interventions cannot be predicted by the model, but when the new situation is incorporated into the model, the forecasts are again reliable. Comparison with ensemble forecasts for 2022 within the European Covid-19 Forecast Hub showed better performance of our model, which can be explained by a model architecture better adapted to the situation in Slovenia, in particular a refined structure for vaccination, and better parameter tuning enabled by the more comprehensive data for Slovenia. Our model proved to be flexible, agile, and, despite the limitations of its compartmental structure, heterogeneous enough to provide reasonable and prompt short-term forecasts and possible scenarios for various public health strategies. The model has been fully operational on a daily basis since April 2020, served as one of the models for decision-making during the COVID-19 epidemic in Slovenia, and is part of the European Covid-19 Forecast Hub.

GOFLMM.jl: Goodness-of-fit for linear mixed models

Jakob Peterlin

University of Ljubljana, Ljubljana, Slovenia

`jakob.peterlin@mf.uni-lj.si`

Linear mixed models (LMMs) are a popular and powerful tool for analyzing clustered or repeated observations for numeric outcomes. LMMs consist of fixed and random components specified in the model through their respective design matrices. Checking if the two design matrices are correctly specified is crucial since misspecifying them can affect the validity and efficiency of the analysis. Together with Rok Blagus and Nataša Kejžar, we figured out how to use specific random processes to test the appropriateness of the assumed design matrices. Furthermore, we show how these processes can be used to test for the goodness of fit of the entire model or its fixed or random component. We have proved that our approach works asymptotically with the help of the theory of empirical stochastic processes and on smaller samples with the help of simulations. We are currently in the process of publishing our work. I will present the package GOFLMM.jl that implements our method in Julia and briefly talk about the wrapper in R that will allow R users to use this quite performant package directly from R.

A simulation study of the limitations of Bland Altman analysis

Maša Kušar

University of Ljubljana, Ljubljana, Slovenia
masa.kusar@mf.uni-lj.si

Bland Altman is a classical and relatively simple method for evaluating whether two measurement methods can be used interchangeably. However, as seen in literature, issues can arise when the method is used in practice. Many of these are demonstrated in the evaluation of various methods of measuring intraabdominal pressure, which we use as our real life example. It can be shown that Bland Altman analysis is unsuitable for comparisons of methods, where the distribution of mistakes from the true values varies greatly between both compared methods, especially cases, where one of the methods can be assumed to give true/very precise values. In such cases, there is a (near) linear relationship between the differences and means of measurement pairs, when differences should ideally be uniformly distributed at all means. The utility of the Bland Altman analysis is also known to be limited in cases where the reference method itself shows poor repeatability. Lastly, the Bland Altman analysis should be adjusted when there are pairs of repeated measurements in the dataset, as is common in medical literature, including our motivational example. In our work, we study the size of the effect of these issues on the validity of conclusions of Bland Altman analysis using simulations. Particularly in the case of hierarchical data, even a relatively small between-subject component of variance leads to a significant error in estimating the limits of agreement.

Measure of expected goals in football and its application

Urh Peček

University of Ljubljana, Ljubljana, Slovenia
pecek.urh@gmail.com

Data analytics has become an important part of serious sports organizations over the past two decades. With the development of sports analytics, experts have developed various statistics and indicators that allow the quantification of certain observations. At the forefront of the measures used in football is the measure of expected goals, denoted by xG (expected Goals). The measure of expected goals represents the probability that a precisely determined shot will be converted into a goal. The concept and idea of calculation of expected goals is presented on the basis of a football game. By simulating a match based on xG values of individual shots we can try to put into context the probability of an outcome of an individual match and show how the match would have played out without the presence of chance. Monte Carlo method to simulate a single match is described. The impact of the distribution of expected goals and its impact on the team's probability of winning a match is presented. Two methods for predicting the results of football games based on the Poisson distribution are presented and updated to take expected goals into account. A practical example with some newly created measures for comparing models is used to compare the methods with and without taking expected goals into account.

Bayesian estimation for linear regression model with multicollinearity problem using Gibbs sampling: A simulation study

Monthira Duangsaphon, Teerawat Simmachan, Kamon Budsaba
and Rattana Lerdsuwansri

Thammasat University, Pathum Thani, Thailand

monthira.stat@gmail.com, teerawat@mathstat.sci.tu.ac.th, kamon@mathstat.sci.tu.ac.th,
rattana@mathstat.sci.tu.ac.th

Multicollinearity is a common problem in multiple regression analysis. The ordinary least squares method provides high standard errors. This results in inefficient hypothesis testing and estimation on the regression coefficient parameters. To overcome this problem, we proposed two schemes of Bayesian estimation with informative prior distributions along with ridge regression as an alternative method. The first scheme, we examined the prior distributions for coefficients and inverse of variance as multivariate normal and gamma distributions, respectively. The second scheme, we considered the prior distributions for coefficients, inverse of variance, and ridge parameter as multivariate normal, gamma and gamma distributions, respectively. The likelihood function and prior distributions lead to the corresponding marginal posterior distribution of the specific parameters. On the basis of the proposed schemes, Gibbs sampling technique was employed to deal with difficulty of obtaining marginal posterior distributions analytically. This technique generates estimators from the conditional posterior distribution of one parameter while fixing the value of the others. Monte Carlo simulation study was conducted to evaluate the performance of the proposed methods compared with Ridge regression and Ordinary Least Squares methods in terms of the total of mean square errors and total of mean absolute bias. The results show that in most cases the second scheme presents the best performance. Moreover, the performance of the first scheme and the Ridge regression are quite similar.

Determining the changepoints in segmented regression

Ana Radulović¹ and Rok Blagus²

¹Institute of Public Health of Montenegro, Podgorica, Montenegro

²University of Ljubljana, Ljubljana, Slovenia

radulovicana@gmail.com, rok.blagus@mf.uni-lj.si

We are interested in estimating the position of change points in segmented regression. Extensive Monte-Carlo simulations are performed to compare different methods: grid search and segmented. In the simulation study, we investigate factors such as sample size, number of changepoints, the position of the changepoints, and the magnitude of the changes in trend between the change points. The point estimates are evaluated and compared via standard estimation bias and precision measures. Coverage of the interval estimators is also evaluated. Finally, the methods are illustrated in a real data example. Comparison of different approaches for determining joinpoints in segmented regression enables researchers to use the currently best available tool for analyzing, understanding, and evaluating these types of non-linear associations and to understand the potential limitations of these methods avoiding potentially flawed conclusions based on erroneous analyses.

Network analysis

Projections of weighted two-mode networks

Vladimir Batagelj

Institute of Mathematics, Physics and Mechanics / University of Primorska, Andrej Marušič
Institute / National Research University Higher School of Economics, Ljubljana/Koper
(Capodistria)/Moscow, Slovenia/Slovenia/Russia
vladimir.batagelj@fmf.uni-lj.si

In a *two-mode* (affiliation or bipartite) network $N = ((U, V), L, w)$ the set of nodes is split into two disjoint sets (*modes*) U and V . Each link $e \in L$ has one end-node in the set U and the other end-node in the set V . The function $w : L \rightarrow \mathbb{R}$ assigns to each link its weight. The network N can be described by the corresponding matrix UV . $UV[u, v] = w(u, v)$ for $(u, v) \in L$, and $UV[u, v] = 0$ otherwise. An approach to the analysis of a two-mode network is its conversion or *projection* to an ordinary (one-mode, weighted) network on a selected mode. This network can be analyzed further using standard network analysis methods. The standard projection to the second mode V is obtained by multiplying the transposed network matrix with the network matrix, $VV = UV^T \cdot UV$. There are some problems with the standard projection (Batagelj, Scientometrics, 2020, 123(2), 621–633). They can be resolved using network normalizations—the *fractional approach*. Especially we point to the role of the first mode nodes of degree 0 or 1. Another type of projection is based on a (dis)similarity measure d on vectors over \mathbb{R} , $VV[v_1, v_2] = d(UV[., v_1], UV[., v_2])$. In many cases, we can show how these measures are related to the standard projection. For illustrations, we present the results of applications of projections to some real-life networks.

Scaling limits for parking on Frozen Erdős-Rényi Cayley trees with heavy tails

Andrej Srakar

Institute for Economic Research (IER) and University of Ljubljana, Ljubljana, Slovenia
srakara@ier.si

In a recent contribution, Contat and Curien (2021) studied parking problem on uniform rooted Cayley tree with n vertices and m cars arriving sequentially, independently, and uniformly on its vertices. In a previous contribution, Lackner and Panholzer (2016) established a phase transition for this process for certain values of m and n . Contat and Curien couple this model with a variant of the classical Erdős-Rényi random graph network process which enables describing the phase transition for the size of the components of parked cars using a (“frozen”) modification of the multiplicative coalescent. Contat and Curien show the scaling limit convergence towards the growth-fragmentation trees canonically associated to the $3/2$ -stable process that appeared previously in the study of random planar maps (Zolotarev, 1986). Yet, their scaling limits unraveled are common to models as long as the degree distribution and the car arrivals have a sufficiently light tail. We study their novel model in the presence of group arrival of cars with heavy tail, and derive the appropriate metric space scaling limits, following Conchon-Kerjan and Goldschmidt (2020), Bhamidi, van der Hofstad and Sen (2018) and Broutin, Duquesne and Wang (2018), with comparing the behaviour of the extended tree parking approach to more commonly studied Bienaymé-Galton-Watson trees. In an application we use The Car Parking Lot Dataset containing 90 000 cars from 4 different parking lots collected by means of a drone, to study the validity of derived limits on a real dataset.

Stochastic generalized blockmodeling

Aleš Žiberna

University of Ljubljana, Ljubljana, Slovenia
ales.ziberna@fdv.uni-lj.si

The aim of this talk is to introduce stochastic generalized blockmodeling for binary networks that is generalized blockmodeling with elements of a stochastic blockmodeling. The approach mainly boils down to using - log-likelihood instead of a measure of variability within homogeneity generalized blockmodeling and with some additional “tweaks”. Such approach might have some of the benefits from both approaches that is the flexibility or adaptability of generalized blockmodeling and the theoretical basis of stochastic blockmodeling. The use of such approach for blockmodeling linked networks will also be discussed.

Making blockmodeling easier with BlockmodelingGUI: R Shiny as a GUI for generalised block-modelling

Fabio Ashtar Telarico

University of Ljubljana, Ljubljana, Slovenia
fabio-ashtar.telarico@fdv.uni-lj.si

This presentation offers a look at an expansion package for the R language for statistical computing available on the CRAN that provides a GUI built in R Shiny to help users access specific functionalities for blockmodeling and, secondarily, network analysis. The app focuses mainly on the package blockmodeling by Aleš Žiberna and deploys some of the network-visualisation capabilities offered by the packages network, igraph and visNetwork. Namely, the app aims at providing a polished interface for one-mode generalised blockmodeling of single-relational networks. In other words, it allows to partition a network into clusters based on pattern of ties and contingent on the selection of a definition of equivalence amongst the one offered. Concretely, the presentation explores and justifies the choice made during the design of the app in terms of allowed types of inputs and outputs. Moreover, it will provide an example of the app's analytical use. In conclusion, the presentation both summarises what has been already achieved and highlights possible future developments.

**Authors and research topics presented at the Applied Statistics
conference (2010–2022)**

Polona-Maja Repar, Bor Vratanař and Andrej Kastrin

University of Ljubljana, Ljubljana, Slovenia

polona-maja.repar@mf.uni-lj.si, bor.vratanař@mf.uni-lj.si,
andrej.kastrin@mf.uni-lj.si

The aim of this study was to systematically examine the knowledge landscape of studies presented at applied statistics conferences over the past two decades. We collected all abstracts published in conference proceedings between 2010 and 2022. We (i) used a natural language processing pipeline to extract, clean, and normalize text data from the Author, Title, and Abstract fields, and (ii) created two co-occurrence networks reflecting the relationships between (i) authors and (ii) keywords. Both networks were then characterized at different levels of granularity (static analysis vs. time slice analysis and whole network vs. node-level analysis). The exploratory analysis revealed both author- and time-specific research topics.

Education and social sciences

Brownian excursions and the Kolmogorov statistic

Mihael Perman¹ and Jon A. Wellner²

¹University of Ljubljana/University of Primorska, Ljubljana/Koper (Capodistria), Slovenia

²University of Washington, Seattle, United States

mihael.perman@fmf.uni-lj.si, jaw@stat.washington.edu

The distribution of the Kolmogorov statistic and related statistics can be derived in different ways. The approach using point processes of marked Brownian excursions provides an elegant derivation of many well distributions in a unified way. The advantage of the approach is that it gives some insight into the structure and interconnection of the various distributions.

Do graduate students of business and management view statistics and research skills as relevant for their careers?

Irena Ograjenšek¹ and Iddo Gal²

¹University of Ljubljana, Ljubljana, Slovenia

²University of Haifa, Haifa, Israel

irena.ograjensek@ef.uni-lj.si, iddo@research.haifa.ac.il

Business and management students often perceive research methods courses as uninteresting and irrelevant (Edwards & Thatcher, 2004). They also often do not know how to access scientific work; have only a limited understanding of scientific methods, and restricted expertise in gaining evidence-based knowledge. This contradicts the increasing focus on data and quantitative thinking in business and management. Educators should therefore strive to prevent a situation where practical knowledge and conventional problem-solving are considered more useful than scientific knowledge, even when the former contradicts scholarly findings (Benson & Blackman, 2003). We report on an innovative study based both on revised and newly proposed measurement scales to capture diverse beliefs and attitudes related to statistics; value of research methods; legitimacy of qualitative and quantitative research in the business and management domain; importance of both quantitative and qualitative reasoning; and more. While most studies to date focus on students of introductory statistics service courses, we further innovate by capturing beliefs and attitudes of graduate business and management students with varying degrees of work experience, who have already entered professional and leadership positions in the labor market or will do so upon their graduation. Preliminary results point to various gaps and interesting patterns. They indicate an intriguing gender difference, i.e., male students have a more positive view of quantitative research for their career. The results also show that students do not necessarily view quantitative methods as more relevant for their future career when they take more quantitative research courses, contrary to expectations that students will develop positive views of statistics and research as they improve their knowledge in this regard. The findings have many implications not only for those teaching statistics and research methods, but also for designers of study and training programs developing future corporate leaders, and point to a new research agenda.

Science literacy and enjoyment in learning science: A statistical perspective

Melita Hajdinjak

University of Ljubljana, Ljubljana, Slovenia
melita.hajdinjak@fe.uni-lj.si

The countries included in the OECD Programme for International Student Assessment (PISA 2015) are compared according to their achievements in science literacy and the index of enjoyment in learning science at the same time, and they are automatically classified into clusters/groups using cluster analysis of patterns. Three clusters of countries are obtained: (i) above-average literate children with below-average enjoyment of learning science (e.g., Japan, South Korea, Slovenia), (ii) above-average literate children with above-average enjoyment of learning science (e.g., China, Singapore, Canada) and (iii) below-average literate children with above-average enjoyment of learning science (e.g., Bulgaria, Turkey, North Macedonia). Interestingly, one would expect a fourth cluster of countries, namely below-average literate children with below-average enjoyment of learning science, but there is no such cluster. The same insight into educational data is obtained from appropriately chosen types of data presentation. While PISA cannot identify cause-and-effect relationships between policies/practices and student outcomes, it can show educators, policy makers and the interested public how education systems are similar and different—and what that means for students. In our particular case, we show how simple statistical analysis or even just a different visualisation of the data can lead to new surprising discoveries. Namely, despite extensive analyses, the finding that none of the countries included in the PISA 2015 survey does have below-average literate children who enjoy learning science below average (as measured through students' responses to the PISA background questionnaire) has never been exposed before and is definitely worth further research.

Online survey panels in modern social science research: Can non-probability sampling replace probability sampling?

Gregor Čehovin and Vasja Vehovar

University of Ljubljana, Ljubljana, Slovenia

gregor.cephovin@fdv.uni-lj.si, vasja.vehovar@fdv.uni-lj.si

Due to the rising costs of collecting survey data, the corresponding research in the social sciences is increasingly conducted online. However, when studying the general population, this often means using online panels whose members initially agree—in exchange for certain incentives—to participate regularly in various surveys. Otherwise, ad hoc recruitment is very costly, as it can only be conducted in the traditional way (in person, by phone, by mail), because there is no sampling frame for e-mail addresses of the general population. Most market and opinion research, as well as a significant portion of academic, government and non-profit research have already switched to non-probability online panels, where participants are selected with various non-probability methods, such as convenience sampling or online self-recruitment. Consequently, these panels are much cheaper compared to online panels based on probability sampling. Nevertheless, the latter in principle ensure much higher data quality, as the positive probabilities of inclusion are known in advance for all units in the population, which is the essential precondition for any statistical inference. Researchers are therefore increasingly faced with an important practical dilemma: Which type of online panels should be chosen, a probability or a non-probability panel? We provide an overview of the meta-studies that address this dilemma and illustrate it with three examples from recent social science research in Slovenia.

Why logistic regression is not always suitable for risk ratio estimation

Jon Pustavrh-Mičović and Nikolaj Candellari

University of Ljubljana, Ljubljana, Slovenia

jp5104@student.uni-lj.si, nc2738@student.uni-lj.si

In practice, we are interested in the connections between whether the patients had unwanted consequences because of disease and whether these are related to any of the predictive variables. The most popular measures for such calculations are the odds ratio and the risk ratio. Although the risk ratio is a much more interpretable quantity, it has the disadvantage of being much more difficult to estimate, while the odds ratio can be easily obtained through logistic regression. The problem is that in practice it often happens that the odds ratio of prospects is calculated and then interpreted as risk ratio. This usually leads to wrong conclusions. The purpose of this seminar is to find out when it is appropriate to evaluate the risk ratio with the odds ratio of prospects and to learn and test two methods by which we evaluate the risk ratio.

Wage gap between men and women in the Israeli economy

Tal Shahor and Javier Simonovich

The Max Stern Yezreel Valley College, Emek Yezreel, Israel
tals@yvc.ac.il, javiers@yvc.ac.il

In Israel, as in many other countries, wage received by women are lower than those received by men. The purpose of this study is to examine which part of the gender wage gap in Israel can be explained by differences in demographic attributes, and preferences of men versus women, and which part cannot be explained, that is, indicates gender discrimination. The study uses the Mincer equation and the Oaxaca decomposition, and utilizes data from a survey carried out by the Israel Central Bureau of Statistics. This study addresses the period 2005 to 2017, which allows examination of existing trends in this field. Results of the study show that during this period there was no systematic change in the unexplained wage gap, which indicates the existence of gender discrimination, even though the gap between men's and women's wages increased towards the end of the period.

Invited lecture

Applications of multiple imputation, from calibrating changing occupation/industry coding systems in the US Census, to evaluating the impact of nonresponse in the Slovenian Plebiscite, to understanding placebo effects in pharmaceutical experiments

Donald B. Rubin

Tsingua University/Temple University/Harvard University, Beijing/Philadelphia (PA)/Cambridge (MA),
China/United States

rubin@stat.harvard.edu

Multiple Imputation (MI) for missing data, first proposed in Rubin (1978), has generated a relatively large corpus of theoretical justification and investigation, some of the earlier work summarized in Rubin (1986, 2006). Despite this rich collection of theoretical work, the main focus of MI has always been on its utility for applications. In this presentation we review two important applications of MI from the past, and preview an important application from the future, which involves MI's use to help disentangle "placebo effects" from "real effects" in double-blind randomized trials of drugs.

Workshop

Generalized blockmodeling in R using blockmodeling package

Aleš Žiberna and Marjan Cugmas

University of Ljubljana, Ljubljana, Slovenia

ales.ziberna@fdv.uni-lj.si, marjan.cugmas@fdv.uni-lj.si

The workshop will cover generalized blockmodeling (Doreian et al., 2005; Žiberna, 2007) of mainly one-mode binary and valued networks in R using “blockmodeling” package (Žiberna, 2021). Only basic knowledge of R and networks/graphs is required. The workshop will cover matrix representation of the network, plotting of such matrices, and of course, clustering the units in the network, that is blockmodeling. Clustering units based on structural, regular and generalized equivalence will be covered. The later implies that also pre-specified blockmodeling will be covered. All aspects of blockmodeling with the blockmodeling package from preparing the data through calling the optimization function (including setting appropriate parameters) to plotting and interpreting the results will be covered. In case of sufficient time and expressed interest, blockmodeling two-mode, multilevel, and linked networks can also discussed.

INDEX

Index of Authors

A

Andersen, PK, 18

B

Batagelj, V, 39

Blagus, R, 38

Brizzi, M, 23

Brusa, L, 29

Budsaba, K, 37

C

Candellari, N, 48

Capriotti, F, 25

Chabert-Liddell, S, 30

Cugmas, M, 31, 51

Č

Čehovin, G, 47

D

Došenović Bonča, P, 27

Duangraphon, M, 37

E

Erčulj, V, 28

F

Ferrari, C, 25

Filzmoser, P, 32

Fošnarič, M, 33

G

Gal, I, 45

Gavrić, D, 27

H

Hajdinjak, M, 46

J

Jamnik, H, 19

K

Kamenšek, T, 33

Kastrin, A, 43

Kežar, N, 24

Košuta, T, 21

Kušar, M, 35

L

Lerdsuwansri, R, 37

Lotrič Dolinar, A, 27

Lusa, L, 20

M

Mandrekar, J, 17

Manevski, D, 21

Marinelli, F, 23

Matias, C, 29

Morgan, T, 27

O

Ograjensek, I, 45

P

Passerini, M, 25

Peček, U, 36

Pedrazzi, G, 25

Perman, M, 44

Peterlin, J, 34

Pohar Perme, M, 18, 21, 22

Pustavrh-Mičović, J, 48

R

Rabatelli, P, 25

Radulović, A, 38

Rahnenfuehrer, J, 20

Repar, P, 43

Rozman, A, 27

Rubin, DB, 50

S

Sambt, J, 27

Schmid, M, 16

Shahor, T, 49

Simmachan, T, 37

Simonovich, J, 49

Sparre Wandall, EN, 18

Srakar, A, 40

Stupica, D, 24

Š

Šarc, I, 27

Šelb, J, 27

T

Tagliaferri, S, 25

Telarico, FA, 42

V

Vehovar, V, 47

Vidmar, G, 19

Vratanar, B, 22, 43

W

Wellner, JA, 44

Z

Ziherl, K, 27

Ž

Žiberna, A, 31, 41, 51

Žibert, J, 33

MY NOTES

